

# Virtualization within FermiGrid

Keith Chadwick



# FermiGrid – The People

---

Keith Chadwick ([chadwick@fnal.gov](mailto:chadwick@fnal.gov))

Chris Green ([green@fnal.gov](mailto:green@fnal.gov))

Neha Sharma ([neha@fnal.gov](mailto:neha@fnal.gov))

Steve Timm ([tim@fnal.gov](mailto:tim@fnal.gov))

Dan Yocum ([yocum@fnal.gov](mailto:yocum@fnal.gov))



## Previous Work

---

Previous talks on “FermiGrid High Availability”

HEPiX 2007 in St. Louis:

<http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2513>

OSG All Hands 2008 at RENC:

<http://indico.fnal.gov/materialDisplay.py?subContId=1&contribId=13&sessionId=0&materialId=slides&confId=1037>

Fermilab detailed documentation:

<http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2590>

<http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2539>



# FermiGrid-HA – Highly Available Grid Services

---

The majority of the services listed in the FermiGrid service catalog are deployed in high availability (HA) configuration that is collectively known as "FermiGrid-HA".

FermiGrid-HA utilizes three key technologies:

- Linux Virtual Server (LVS).
- Xen Hypervisor.
- MySQL Circular Replication.



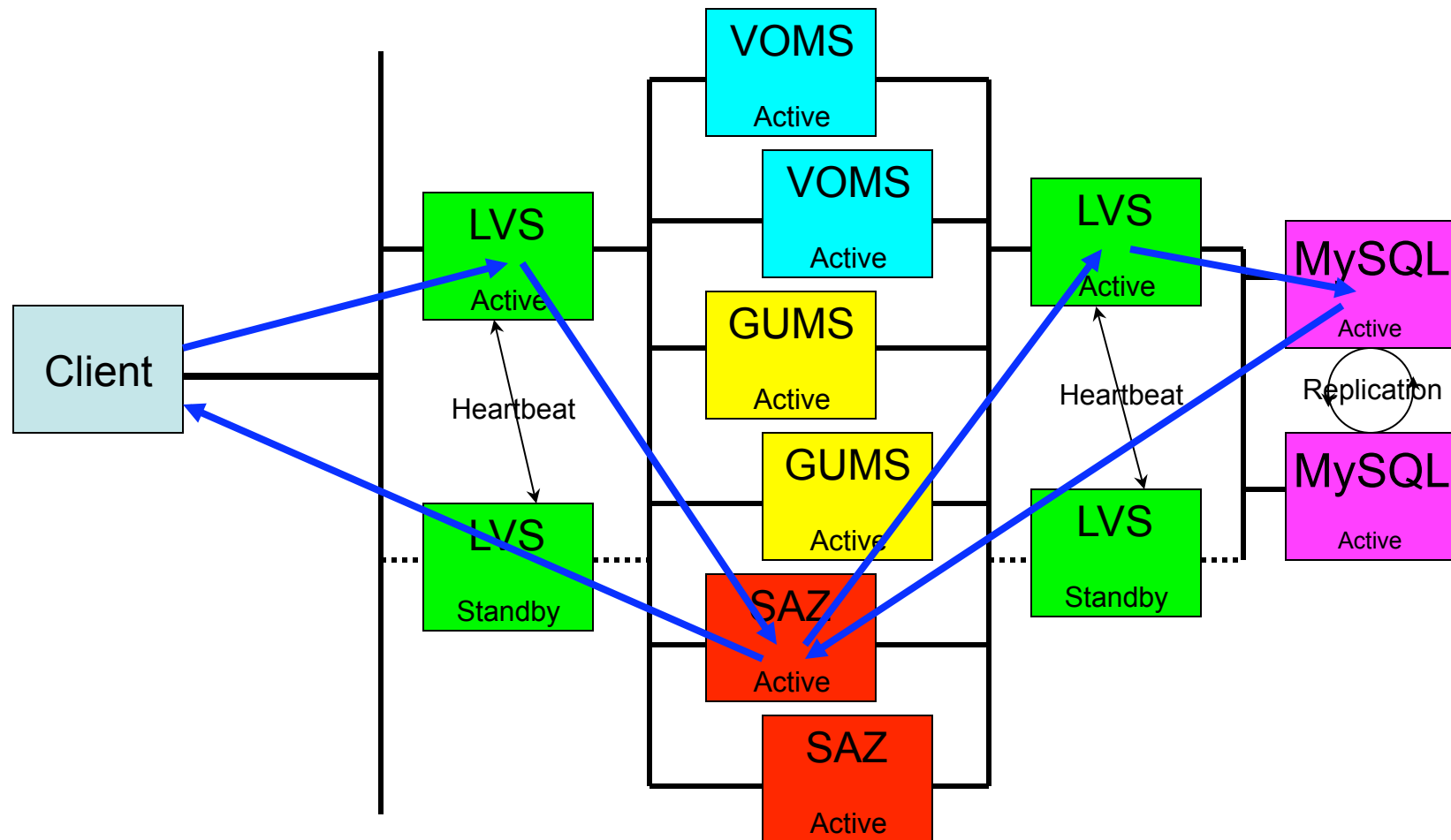
# HA Services Deployment

---

FermiGrid employs several strategies to deploy HA services:

- Trivial monitoring or information services (such as Ganglia and Zabbix) are deployed on two independent virtual machines.
- Services that natively support HA operation (Condor Information Gatherer, FermiGrid internal ReSS deployment) are deployed in the standard service HA configuration on two independent virtual machines.
- Services that maintain intermediate routing information (Linux Virtual Server) are deployed in an active/passive configuration on two independent virtual machines. A periodic heartbeat process is used to perform any necessary service failover.
- Services that do not maintain intermediate context (i.e. are pure request/response services such as GUMS and SAZ) are deployed using a Linux Virtual Server (LVS) front end to active/active servers on two independent virtual machines.
- Services that support active-active database functions (circularly replicating MySQL servers) are deployed on two independent virtual machines.

# HA Services Communication





# FermiGrid – Organization of Physical Hardware and Virtual Services

---

<http://fermigrid.fnal.gov/fermigrid-systems-services.html>

- <http://fermigrid.fnal.gov/fermigrid-organization.html>
- <http://fermigrid.fnal.gov/cdfgrid-organization.html>
- <http://fermigrid.fnal.gov/d0grid-organization.html>
- <http://fermigrid.fnal.gov/gpgrid-organization.html>
- <http://fermigrid.fnal.gov/gratia-organization.html>
- <http://fermigrid.fnal.gov/fgtest-organization.html>



# Virtualized Non-HA Services

---

The following services are virtualized, but not currently implemented as HA services:

- Globus gatekeeper services (such as the CDF and D0 experiment globus gatekeeper services) are deployed in segmented pools.
  - Loss of any single pool will reduce the available resources by approximately 50%.
- MyProxy
- OSG Gratia Accounting service [Gratia]
  - not currently implemented as an HA service.
  - If the service fails, then the service will not be available until appropriate manual intervention is performed to restart the service.
- OSG Resource Selection Service [ReSS]
  - not currently implemented as an HA service.
  - If the service fails, then the service will not be available until appropriate manual intervention is performed to restart the service.

We are working to address these services as part of the FermiGrid FY2009 activities.





# Measured Service Availability

---

FermiGrid actively measures the service availability of the services in the FermiGrid service catalog:

- <http://fermigrid.fnal.gov/fermigrid-metrics.html>
- <http://fermigrid.fnal.gov/monitor/fermigrid-metrics-report.html>

The above URLs are updated on an hourly basis.

The goal for FermiGrid-HA is > 99.999% service availability.

- Not including Building or Network failures.
- These will be addressed by FermiGrid-RS (redundant services) in FY2010/11.

For the period 01-Dec-2007 through 30-Jun-2008, we achieved a service availability of 99.9969%.

For the period 01-Jul-2008 through the present, we have achieved a service availability of 99.9813% (and climbing...).



# FermiGrid Service Level Agreement

---

## Authentication and Authorization Services:

- The service availability goal for the critical Grid authorization and authentication services provided by the FermiGrid Services Group shall be 99.9% (measured on a weekly basis) for the periods that any supported experiment is actively involved in data collection and 99% overall.

## Incident Response:

- FermiGrid has deployed an extensive automated service monitoring and verification infrastructure that is capable of automatically restarting failed (or about to fail) services as well as performing notification to a limited pager rotation.
- It is expected that the person that receives an incident notification shall attempt to respond to the incident within 15 minutes if the notification occurs during standard business hours (Monday through Friday 8:00 through 17:00), and within 1 (one) hour for all other times, providing that this response interval does not create a hazard.

## FermiGrid SLA Document:

- <http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=2903>



# Why 99.999%?

---

A service availability of 99.999% corresponds to 5m 15s of downtime in a year.

- This is a challenging availability goal.
- [http://en.wikipedia.org/wiki/High\\_availability](http://en.wikipedia.org/wiki/High_availability)

The SLA only requires 99.9% service availability = 8.76 hours.

So, really - Why target five 9's?

- Well if we try for five 9's, and miss then we are likely to hit a target that is better than the SLA.
- The hardware has shown that it is capable of supporting this goal.
- The software is also capable of meeting this goal (modulo denial of service attacks from some members of the user community...).
- The critical key is to carefully plan the service upgrades and configuration changes.



# FermiGrid Persistent ITB

---

Gatekeepers are Xen VMs.

Worker nodes are also partitioned with Xen VMs:

- Condor
- PBS
- Sun Grid Engine
- + A couple of "extra" CPUs for future cloud investigation work.

<http://fermigrid.fnal.gov/fgitb-organization.html>



# Cloud Computing

---

FermiGrid is also looking at Cloud Computing.

We have a proposal in this FY, that if funded, will allow us to deploy an initial cloud computing capability:

- Dynamic provisioning of computing resources for test, development and integration efforts.
- Allow the retirement of several racks of out of warranty systems.
  - Both of the above help improve the “green-ness” of the computing facility
- Additional capacity for the GP Grid cluster.

## Virtualization is working well within FermiGrid.

- All services are deployed in Xen virtual machines.
- The majority of the services are also deployed in a variety of high availability configurations.

## We are actively working on:

- The configuration modifications necessary to deploy the non-HA services as HA services.
- The necessary foundation work to allow us to move forward with a cloud computing initiative (if funded).

Any questions?